

Random Forests

Justin Dixon

2018-02-28T21:13:14-05:00

Introduction

Following on from the previous post about decision trees let us move on to Random Forests. Let us use the Soybean data from the ‘mlbench’ package. There are 35 features and 683 observations with 16 varieties of Soybean.

Why care about Random Forests?

Let us look at how our decision trees predict previous unseen data. First we will load the data in:

```
library(mlbench)
library(caret)
data("BreastCancer")
dim(BreastCancer)
```

Let us now split the data up into a training and test data set.

```
BreastCancer <- na.omit(BreastCancer)
BreastCancer$Id <- as.factor(BreastCancer$Id)
BreastCancer <- subset(BreastCancer, select = -Id)
index <- createDataPartition(y = BreastCancer$Class, times = 1, p = 0.75, list=FALSE)
train0 <- BreastCancer[index,]
test0 <- BreastCancer[-index,]
```

We can now train a decision tree like we did in the previous tutorial and see how the model did predicting values of the test set.

```
Tree <- caret::train(Class ~., data = train0, method = 'rpart')
#Tree <- rpart(Class ~., data = train0)
results <- caret::predict.train(Tree, newdata = test0)
hello <- caret::confusionMatrix(results, test0$Class)
overall <- hello$overall
print(knitr::kable(t(overall)))
```

```
print(knitr::kable(hello$table))
```

That is a pretty good model but lets see if we can improve it. Decision trees are know to have lower bias and this equates to a higher variance. That is decision trees *overfit* the training data. Random forests allow us to overcome this issue.

```
RF <- caret::train(Class ~ ., data=train0, method = 'rf', ntree= 300)
results <- caret::predict.train(RF, newdata = test0)
hello1 <- caret::confusionMatrix(results, test0$Class)
overall1 <- hello$overall
print(knitr::kable(t(overall1)))
```

```
print(knitr::kable(hello1$table))
```

See, we achieved a slightly higher accuracy. That means we successfully classified more cancer patients. That is a pretty good thing, well worth using random forests over a single decision tree.

The maths